DOCUMENT RESUME

ED 189 175 TM 800 357

AUTHOR Ellett, Chad D.
TITLE Assessing Minimum Competencies of Beginning Teachers: Instrumentation, Measurement Issues, Legal Concerns.
PUB DATE Apr 80
NOTE 34p.: Paper presented at the Annual Meeting of the National Council on Measurement in Education (Boston, MA, April 8-10, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Beginning Teachers; *Competency Based Teacher Education; Higher Education; Legal Problems; *Measurement Techniques; Standards; *Teacher Certification; Teaching Skills; Test Construction; Testing Problems; Test Reliability; Test Validity
IDENTIFIERS Georgia; *Performance Based Certification; *Teacher Performance Assessment Instruments

ABSTRACT

An overview is presented of a performance-based assessment system, Teacher Performance Assessment Instruments (TPAI), developed by the Teacher Assessment Project at the University of Georgia to measure competencies of beginning teachers for initial professional certification. To clearly separate the preparation and certification functions within the state, the Georgia Department of Education created a Performance Based Certification (PBC) model for prospective professional teachers. Beginning in the fall of 1980, the subject field proficiency of all beginning teachers will be demonstrated through the use of a criterion-referenced knowledge test. When employed as a beginning teacher, the TPAI will be used for assessment. Five performance assessment instruments have been developed: (1) Teaching Plans and Materials; (2) Classroom Procedures; (3) Interpersonal Skills; (4) Professional Standards; and (5) Student Perceptions. These instruments are admininstered through observation and/or interview with the teacher being assessed. Georgia is the first state to commit itself to the implementation of a PBC model for beginning teachers. Some developmental procedures and problems, measurement problems, performance standards, and the legality of performance assessment systems for certification are discussed. Other associated issues are presented in consideration of many southeastern states presently contemplating similar PBC models. (Author/GSK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED189175

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Assessing Minimum Competencies of Beginning Teachers: Instrumentation, Measurement Issues, Legal Concerns

Chad D. Ellett
University of Georgia

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
C. Ellett

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the Annual Meeting of the
National Council on Measurement in Education
Boston, Massachusetts
April 8, 1980

TM 800357

Introduction

The movement toward competency based education (CBE) and performance based certification (PBC) for public school teachers is a rapidly expanding phenomenon in American education. The CBE model and its application to pre-service and in-service teachers is growing at the same time that the majority of states have either considered or are actually implementing minimum competency testing for students.

Since 1967, when Congress passed the Education Professions Development Act which was designed to prepare and maintain an adequate supply of competent teachers, the CBE concept has rapidly expanded as a philosophical base for program development and state certification and training models. According to an early report by Schmeider (1973), more than 500 teacher education programs claimed to be competency based or were in the developmental stages of moving in that direction.

As with most "innovations" in education, the CBE model, whether applied to students, teachers, administrators, or others, is fraught with many complex issues and problems. Some of these are philosophical, some legal, some simply logistical, and others of a more traditional research, development, and measurement nature. All of these issues and problems have to be adequately addressed and resolved if the CBE movement is to become something other than an educational "fad" and if CBE is to remain a viable force in changing American education.

This paper attempts to address some of the more important issues surrounding the Georgia State Department of Education's commitment to implement the CBE/PBC model with beginning and student teachers during the 1980-1981 school

year. The issues are derived from experiences in developing assessment instruments to measure minimum teaching performance on a set of competencies developed through the Teacher Assessment Project at the University of Georgia. No attempt will be made to discuss the basic assumptions, requirements, and operatives of a CBE model since excellent expositions of these can be found elsewhere (Houston and Howsam, 1972; Burke, Hansen, Houston, and Johnson, 1976).

The Georgia PBC Plan

For the past few years, the Georgia State Department of Education has rapidly moved toward the implementation of a PBC model for all beginning and student teachers in the state. This movement has received rather widespread support from various publics with vested interests in the training and certification of teachers, e.g., state legislators, professional education organizations, state board of education, and LEAs. At the heart of the movement is the "message" that teachers should be trained around, and expected to demonstrate acceptable performance on a set of competencies considered as minimally essential for successful teaching, before application for permanent certification can be made. The implementation of this vast plan has certainly not been without its many problems, but it does seem to be better received by all interested parties than earlier efforts in other states, e.g., California, Texas, and New York. The reasons for success of the Georgia PBC model are not easily identified. However, the ongoing commitment to supportive supervision and staff development activities for teachers, and the desire for a clear separation between the State Department of Education and Higher Education Institution responsibilities for preparation and certification has undoubtedly contributed to the successful implementation of the model. In addition, a large scale research and development effort has been underway

at the University of Georgia to development valid and reliable instrumentation to assess both beginning and student teacher performance in the classroom setting.

The current certification model in Georgia, as in most states, depends heavily upon recommendations for certification to the State Department of Education from the various colleges and universities in the state having approved teacher education programs. These recommendations have traditionally been made using a variety of criteria, most of which have no demonstrated relationship to actual classroom performance, e.g., GPA, course offerings completed, NTE scores, student teaching supervisor ratings, etc. If the higher education institution has a state approved teacher education program, then it essentially functions as the "certifier" in the sense that students completing program requirements are then recommended for certification. In most all instances, the recommendations for certification are merely "rubber stamped" by the State Department of Education. This approach to certification has served to minimize performance standards in the classroom, and places heavy responsibility for preparation and certification on the teacher training institution. Thus, deciding whether or not a beginning teacher should be certified has been the primary responsibility of the institution of higher learning where the person obtained his/her teacher education degree.

The Georgia Department of Education, in an effort to create a better balance among the various criteria for use in certifying beginning teachers, and to more clearly separate the preparation and certification functions within the state, is creating and implementing a PBC model for prospective professional teachers. The current plan includes a performance assessment of prospective professional teachers at two key points: (1) during student teaching, and (2) during the initial years of the on-the-job experience.

Beginning in the Fall of 1980, all beginning teachers will demonstrate proficiency in their subject matter area (certification field) by meeting minimal performance standards on a criterion-referenced knowledge test. Thirty two of these tests are currently being administered. If the student teacher satisfies the requirements for the years of preparation and passes a criterion-referenced test in the teaching field, he/she is provided with a non-renewable certificate valid for three years. When employed as a beginning teacher, the teacher will be assessed with the Teacher Performance Assessment Instruments(TPAI) developed by the Teacher Assessment Project (TAP) at the University of Georgia. When minimally essential skills for teaching are demonstrated, the beginning teacher will be issued a renewable, professional teaching certificate. The current PBC plan includes allowing the beginning teacher three years and a minimum of six assessments to demonstrate proficiency on each teaching competency (skill) assessed by the TPAI.

In response to the State Department of Education's mandate to move toward a PBC model, the Teacher Assessment Project (TAP) was funded to do the research and development work necessary to identify performance competencies to be used in the PBC plan, and to develop and field test instrumentation to facilitate implementation of the model. The sections of this paper that follow summarize these research and development efforts to date.

Developing and Verifying the Competencies

Competency identification for the TPAI began with a review of numerous studies conducted in Georgia and other areas of the nation to determine which competencies were appropriate for consideration for beginning teacher performance assessment. The primary studies used in this literature search were three large-scale, state-supported investigations conducted between 1972 and 1975. In addition, research findings emerging from teacher effectiveness

studies were incorporated into the search for suitable and meaningful competency statements. These reviews resulted in a total number of 219 competencies derived from available sources. These original statements were consolidated so as to eliminate apparent redundancy in content. They were then edited into a consistent and meaningful style which resulted in a revised list of 52 statements.

After review by Department of Education personnel and an Ad Hoc Committee of practitioners and theorists in teacher education, the competencies became content for a survey instrument. The instrument was designed to collect opinions about how important the competencies were for the enhancement of pupil learning in the classroom, for beginning teachers to possess, and for teachers seeking professional status (renewable certification). The instrument was administered to a representative sample of education "interests" throughout the state in 153 public schools and 31 colleges and universities. Besides rating items, participants were asked to make comments and suggestions about the competencies. The total number of participants was 4,668, which was found to be representative of Georgia's educators as regards geographical distribution, census classification, and professional roles.

When all responses to the survey instrument had been summarized, the results suggested that 33 of the 52 competencies were rated as <u>essential</u> for teaching, rather than highly desirable or desirable. The same competencies were viewed by the survey participants as essential for pupil learning, initial teaching, and for continuous professional performance. Levels of behavior found to be necessary for professional performance were predictably higher than those deemed necessary for initial teaching performance. In addition, there was overwhelming agreement among different subgroups in the sample, i.e., administrators, college faculty, school teachers, as to which competency state-

ments were most important, as well as to the order of their importance. A complete report of the development and verification of the TPAI competencies can be found in Johnson, Okey, Capie, Ellett, and Adams (1978). The most significant outcome of these efforts was the identification of a set of "generic" competencies (performance statements) that formed the framework for the development of the ensuing Teacher Performance Assessment Instruments (TPAI).

Development of the Teacher Performance Assessment Instruments (TPAI)

Once conclusions from the statewide competency verification study were evident, the Teacher Assessment Project (TAP) staff began the development of the Teacher Performance Assessment Instruments (TPAI) to be initially field tested in the state during the 1977-78 school year. Because of the nature of the competency statements and other considerations, five performance assessment instruments were developed: (1) Teaching Plans and Materials (TPM), (2) Classroom Procedures (CP), (3) Interpersonal Skills (IS), (4) Professional Standards (PS), and (5) Student Perceptions (SP). Except for Student Perceptions, competencies on the separate instruments encompass several performance indicator statements which are scored by reference to a series of five scale-point descriptors. An example of a TPAI competency statement, indicator, and associated scale descriptors can be found in Appendix A. As with the competency statements, statewide verification surveys were undertaken with a variety of groups of educators using the indicators and descriptor sets to evaluate their appropriateness for teaching, certification, pupil learning, and instrument classification. The 1978-1979 form of the TPAI Instruments for beginning teachers consists of 16 competencies and 51 performance indicators with associated descriptors for each indicator statement. Only 14 competencies and associated

indicators (TPM, CP, and IS Instruments) are to be used for certification. A complete listing of 1979-80 TPAI Competencies and indicators can be found in Appendix B.

TPAI Assessment Procedures

The TPAI instruments are administered through actual observation and/or interview with the beginning teacher being assessed. Four data sources are used in administering the complete instrument battery: (1) a trained, external Data Collector, (2) a peer teacher or colleague nominated by the beginning teacher, (3) a school administrator (usually the principal), and (4) students in the teacher's classroom (grade 3 and above). A significant part of the total TAP project has been the development of an elaborate training program (approximately 50 hours) for data collectors to learn to conduct a reliable and valid assessment with TPAI. This training program has been used to train approximately 5,000 persons throughout the state of Georgia. The state has funded 17 Regional Assessment Centers to Coordinate TPAI data collection, results feedback, and staff development activities.

The general assessment plan calls for the trained data collector to discuss with the beginning teacher how the assessment will be undertaken, when classroom performances will be observed, and what materials, lesson plans, etc., will be needed to complete the assessment. The process is in no way a "secret" one, and the beginning teacher is fully informed about the assessment procedures and purpose of the assessment. A complete assessment of a beginning teacher by all data sources, including actual interview and observation time, requires approximately 5-6 person hours.

In addition to a training program, a data collection and management system has been developed by TAP personnel to process data for beginning teachers

throughout the state of Georgia. A computer printout listing competency and indicator scores is then returned to the beginning teacher (See Appendix C for sample). Currently, such profiles have been used primarily for staff development purposes since the state PBC plan will not become fully operational until the fall of 1980. It is expected that from 6,000 to 12,000 TPAI assessments will be processed when PBC takes place during the 1980-1981 school year.

Field Testing the TPAI

Since 1977, a variety of research studies have been undertaken with the TPAI aimed at identifying their psychometric characteristics. In addition to content validation studies, research has focused on establishing criterion-related validity using multiple school-related criteria and on various forms of interrater agreement and classical reliability analyses. Highlights and general findings from these studies with the TPAI are reviewed below. More detailed summaries of these studies are available from the Teacher Assessment Project, College of Education, University of Georgia.

Reliability

Several kinds of reliability procedures have been undertaken to date with the TPAI instruments. First, internal consistency reliability coefficients were computed on a sample of 208 beginning teachers to assess item homogeneity throughout the instrument components. Cronbach Alpha coefficients for these analyses ranged from a low of approximately .75 to a high of .95 for both the pool of competency scores and the pool of indicator scores. Separate computations on various data sources indicated similar findings across sources (i.e., separate analyses for ratings of administrators, peer teachers, and trained data collectors). These findings suggest that the TPAI

items are generally homogeneous as an item set, and demonstrate acceptable internal consistency characteristics across various data sources.

Inter-rater agreement studies have been undertaken with a variety of different methodologies and samples. Some of these have involved viewing a series of videotaped teaching performances in a variety of different teaching situations, i.e., elementary and secondary math, reading, language arts, etc., by both trained and untrained observers. In addition, inter-rater agreement studies have been undertaken in actual classrooms of beginning teachers. Results of these studies now support between 80 and 90 percent "exact agreement" (raters giving identical score to an observed teacher performance). These findings have generalized across several different videotaped performances using a variety of different observers (data collectors) and in a variety of classroom settings.

One interesting finding emerging from inter-rater agreement studies with the TPAI is the tendency to achieve greater inter-rater agreement in actual (live) classroom situations than with videotaped teaching performances. This agreement, depending upon instrument component, ranges from 10 to 15% higher than agreement using videotyped teaching performance. An interpretation of this finding in terms of the inability of videotaped sequences to provide necessary information for making TPAI judgments is offered. Additional analyses of inter-rater agreement data are currently underway using the length of training and observation as a comparative variable. A summary of recent inter-rater agreement study findings can be found in Capie, Ellett, and Johnson, (1979).

Validity

TPAI assessments have been related to a variety of educational outcomes, including the more traditionally examined one in process/product research . . .

student learning. TPAI assessments on a sample of 30 beginning teachers were correlated with pupil achievement gain scores using the 1970 version of the Metropolitan Achievement Test (MAT) (Okey, Capie, Ellett, and Johnson, 1978). These results were not extremely encouraging since only 4 of the TPAI/MAT correlations were statistically significant. A more recent replication of this study with the 1978 Survey Edition of the MAT in 76 elementary classrooms also yielded few significant TPAI/Pupil Standardized Achievement Gain score correlations.

In addition to standardized tests, TPAI competency validation was explored by correlating pupil achievement gain scores on teacher-made tests over two to three week units of instruction. In an initial study, 12 of 20 TPAI competency scores were found to significantly correlate with student learning gains (Okey, et. al., 1978). A recent replication in 56 classrooms (Ellett, Capie, Johnson, 1980) found that 15 of 16 TPAI competencies significantly and positively correlated with pupil learning gains. These findings are rather encouraging indeed, given the large number of failures in past attempts to relate teacher behavior assessments to pupil learning gains in process/product research. Considered together, these findings indicate that teacher-made tests geared to educational objectives in the classroom may be more sensitive criteria for teacher performance validation studies than standardized achievement measures. This makes intuitive sense because pupil performance on standardized tests probably reflects many out of class learning experiences and other factors over which teachers have no direct control, e.g., quality of home environment, pupil aptitude and ability.

TPAI assessments have also been related to student engaged time. Analysis of the 1978 Spring data from 28 secondary classes generated significant TPAI/"On Task" behavior correlations for 14 of the 16 TPAI competencies, ranging

from a low of approximately .40 to a high of .70 (Okey, et. al., 1978). These findings strongly suggest that a beginning teacher rated highly on the TPAI exists in a classroom where random samples of students are observed to be actively and highly involved (engaged) in the instructional process. These data are encouraging because of the believed importance of pupil engaged time to subsequent learning outcomes. Additional studies of this kind are currently underway with the TPAI.

Measures of pupils' perceptions of school climate/learning environment characteristics have also been used as criteria in the TPAI validation model. In the Okey, et. al. study (1978), the Learning Environment Inventory (LEI) (Walberg and Anderson, 1973) was administered to secondary students in 28 classes. The LEI taps student perceptions of 15 different classroom characteristics such as Apathy, Cohesiveness, Disorganization, Difficulty, Competitiveness, Satisfaction, etc. Of a possible 300 TPAI/LEI subscale correlations, 156 were statistically significant ranging in magnitude from .37 to .77. A more recent study (Ellett, Capie, and Johnson, 1980) has generally replicated these findings for both elementary and secondary classes. These data suggest that a variety of important student perceptions of school climate/learning environment characteristics relate to teacher performance assessments in the classroom. The literature concerning the relationship between such perceptual measures and educational criteria (including pupil achievement) is quite large. Studies with the TPAI, however, may be the first investigations relating teacher performance assessment to student perceptions of the school climate/learning environment. Similar replications of these studies are currently underway.

Concurrent validity characteristics of the TPAI have been investigated by comparing TPAI competency scores for teachers with varying years of teaching experience, and by comparing TPAI assessments with the Teacher Practices

Observational Record (TPOR) and Purdue Observational Rating Scales (PORS) instruments. In the study examining TPAI scores for teachers with varying degrees of teaching experience (from zero to five prior years), results clearly demonstrated that more experienced teachers scored progressively higher on the TPAI competencies than beginning teachers, a finding interpreted as supporting assumptions about the ability of the TPAI to discriminate between the quality of beginning and more experienced teachers' performances. The comparable instruments investigation (correlating TPAI scores with scales of the TPOR and the PORS instruments) indicated little or no relationship between the TPAI and the TPOR; but frequently occurring, significant correlations between the TPAI and the PORS. Twelve of 19 possible TPAI CP indicator/PORS correlations were statistically significant, ranging from .32 to .75. In addition, five of nine possible TPAI IS indicator/PORS correlations were statistically significant, ranging from .37 to .41. (Ellett, Johnson, Capie, and Okey, 1978).

The comparable instruments investigation of the TPAI also provides preliminary evidence for the "construct" validity of the instruments when compared to other instruments designed to assess a characteristic labeled "quality" teaching. Additional construct validity support for the TPAI stems from a series of completed factor analysis studies which identified 4 primary TPAI factors largely described by the originally classified TPAI competencies and indicators (Tobin, Capie, Ellett, and Johnson, 1980). The results of several of these analyses indicate that the TPAI competencies "hang together" rather well in describing major teaching dimensions and replicate across various samples of beginning teachers.

In summary, research investigations of the TPAI instruments during their initial development provide data suggesting that the instruments are adequately

valid and reliable for their intended purposes within the state PBC plan. Additional reliability and validity investigations are being undertaken during the current project year with larger sample sizes, and with attention given to replication across various teacher groups. In addition to these studies, reliability and validity investigations have been undertaken with the TPAI on samples of student teachers (Capie, Ellett, and Johnson, 1980).

The sections of this paper than follow present a discussion of issues and problems associated with the implementation of the Georgia State Department of Education PBC plan for beginning and student teachers. While the description of these issues is derived from experiences in Georgia, they are seemingly relevant for other states, and many cut across the entire CBE movement, including programs aimed at establishing minimum competency examinations for pupils as a requirement for graduation.

Issues and Problems

Measurement Issues

Attempts to develop instrumentation to assess teacher performance for certification purposes like the TPAI are fraught with many measurement issues and problems. If these instruments are to be used for certification purposes, they will have to demonstrate adequate reliability and validity characteristics to withstand legal and professional challenges. The basic measurement issue centers on determining which validity and reliability characteristics are most appropriate for certification purposes, and what levels of these are sufficient to insure instrument quality and the accuracy of certification decision. The comments concerning measurement issues offered below are derived from developmental experiences with the TPAI and the Georgia legislative mandate to implement a PBC model for all beginning teachers in the fall of 1980.

Reliability: In developing an observational/performance rating system such as the TPAI, several issues are central to the concept of reliability. One issue is the degree to which such instruments should possess reliability characteristics as defined by traditional approaches (temporal stability, internal consistency, generalizability analyses) as opposed to a major emphasis on inter-rater agreement.

With the TPAI, multiple data sources (trained data collector, school administrator, and peer teacher) are used to generate competency and competency indicator scores. Ratings from all three sources are combined to arrive at competency and indicator scores. An essential question concerning reliability of ratings (particularly as regards certification decisions) is: Do these data sources agree when they observe a particular teaching performance in the classroom? Since the TPAI assessment process requires classroom observation by data sources at different times, several variables (observer, teacher performance, classroom context) are potential sources of unreliability when reliability is defined as inter-rater agreement. Given the state plan to base a beginning teacher's score on multiple data source ratings for certification, should data sources be "weighted" in the decision-making process based on reliability information. Should any sources be dropped from the assessment if "unreliability" is demonstrated?

Classical test and measurement approaches suggest that adequate test-retest reliability (temporal stability) is an important aspect of any measuring device. If the purpose of the TRAI assessment is to determine whether beginning teachers can perform at a minimally acceptable level of performance, is temporal stability reliability necessary to support the use of the TPAI for certification. Is it necessary to demonstrate that TPAI assessments correlate highly over short periods of time when the state assessment for certification process depends

upon combining data from two consecutive assessments conducted in the Fall and Spring of each year? The TPAI are in a sense "criterion-referenced," however, one demonstration of successful performance on a given competency is not considered reliable enough to "pass the test." Given the large number of contextual factors that can change in the classroom setting (including teacher behaviors from one assessment to the next) and the TPAI system of using multiple data sources and multiple assessments, correlations from temporal stability studies might be predictably lower than those considered acceptable for traditional paper and pencil tests. More important, perhaps, is to demonstrate the ability of rater's scores using videotaped teacher performances. Such studies do not yield complete data for the TPAI that matches an actual assessment, but they do help determine the extent to which error (unreliability) is incurred through observer/rater effects.

Validity: Many classroom process studies have been undertaken during the past 25 years, and these have become an important aspect of research on teaching. Two types of teaching process studies have seemingly evolved from such research: those concerned with describing or defining the teaching process, and those attempting to determine which teaching processes are effective in producing desired educational outcomes, such as student achievement. The latter type of research, termed process-product research (Dunkin and Biddle, 1974), has dominated research on teaching during the past 10 to 15 years. Excellent reviews of this research can be found in other sources, and no attempt to review it will be made here. Suffice it to say that process-product research has met with mixed results, with little replication of results from one study to the next. The process-product paradigm, however, becomes an important aspect of demonstrating the validity of the TPAI

and similar instruments designed to assess teaching performance for certification purposes. What kinds of validity characteristics should the TPAI and similar measures have?

Past court decisions about the legality of performance measures for employee selection, certification, and licensure have well-demonstrated the importance of content validity. Professionals in the field must agree that the items and/or factors being assessed are indeed job-relevant. Indeed, a key component of the Equal Employment Opportunity Commission guidelines under Title VII of the Civil Rights Act was the idea that any test must "measure the person for the job," not the person in the "abstract." But, is content validity enough?

The process-product paradigm and the movement toward educational accountability seemingly dictate that attempts be made to relate TPAI assessments to other important educational variables. The most important variable from an accountability perspective is, of course, pupil learning. Studies of the criterion-related validity of teacher performance instruments using student achievement as a criterion have, as mentioned above, met with mixed success and few meaningful replications. Why?

Standardized achievement test results as a pupil learning criterion present several difficulties. First, performance on such measures probably has far more in common with factors other than teacher classroom behavior. Home environment characteristics, a child's innate ability, learning experiences outside of school, etc., account for far more variance in standardized test performance than effective teaching in the classroom. Similarly, the reliability of gain scores from pre to post testing over short periods of time is suspect on such measures, and the stability of "real" gain is estimated at only .30 to .35 over a two-year period. If standardized achievement tests are

used as a validation criterion for teacher performance measures, few such instruments will ever have adequate validity.

An alternative approach currently being implemented in the TAP is to use teacher-made tests over short units of instruction. Large-scale attempts are made to correlate pupil learning gains on such tests with TPAI assessments. Assessments are made on teachers at the time they are teaching content directly related to learning objectives from which test items for pupils are developed. One issue in this validation process is the potential unreliability of teacher-made tests. Content of the tests is difficult to dictate. However, attempts have been made in research with the TPAI to hold structured workshops with teachers to control the methodology by which tests are developed (e.g., test length, item type, difficulty level, etc.) (Ellett, Capie, and Johnson, 1980).

If one were to view recent court decisions concerned with the use of tests for certification, content validation is again seen as an important process for establishing legal acceptability. The recent supreme court decision which supported the use of the National Teacher Examinations as part of the teacher certification process in the State of South Carolina is a case in point. The NTE and its content was essentially validated in the state by asking key educators if the content was relevant to teaching in the public schools. With a high degree of affirmation among those surveyed and the establishment of a very minimal "cut" score, the Supreme Court upheld the use of the NTE for certification purposes, though most large scale reviews of its "job-relatedness" have meet with very mixed results (e.g., see Quirk, 1973).

Unlike paper and pencil measures such as the NTE, the TPAI are, in a sense, the criteria rather than the predictor of teaching performance. Thus, content validation studies seem the necessary and sufficient process for

establishing the legal basis to support the use of performance measures as part of the certification process. Establishing the predictive validity of the TPAI using school outcome variables (e.g., pupil achievement) as criteria, (while interesting from a teacher effectiveness research standpoint) does not seem necessary to insure the legality of the Georgia PBC plan. If consensus is received from "expert" judges (educators) concerning which performances, and what levels of performance should be established for initial teacher certification, this alone seems legally sufficient. Several of these investigations have already been undertaken with the TPAI (Johnson, Capie, and Ellett, 1978) and more are currently in process.

As regards the development of the TPAI and similar instruments, a final comment concerning reliability and validity seems in order from a classical measurement viewpoint. Unless these psychometric characteristics are at acceptable levels, then the case could be made that such instruments measure more of something other than "competence" than competence itself.

Standards Setting and "Cut-off Scores": The determination of criterion levels of performance for initial teaching certification (minimal standards) presents possibly the most difficult measurement and decision-making problems of all. Some of the associated concerns are of a sociopolitical and practical nature rather than a statistical nature. As has been pointed out elsewhere (Hambleton, Powell, and Eignor, 1979), all methods of standards setting are in essense arbitrary. In setting standards, major issues center on how much faith should be placed in empirical (statistical) decision-making models and how much faith should be placed in informed human judgements about performance.

Two general notions underly most of the current thinking regarding performance (as well as paper and pencil) measures. The first can be called "norm-referenced," in the sense that it represents a Bayesian approach to

establishing cutoff scores (performance levels). This model is probabilistic. Given a demonstrated level of performance, what is the probability that an individual is "competent?" What is the probability that the individual is "not yet competent?" In such models, measurement error becomes an important consideration in making judgements about competence of individuals. This model has been explored in detail by Berk (1976), though it has not been widely applied to certification models for teachers. The statistical, decision-making framework can obviously lead to two important classification errors: (1) persons classified as "competent" that actually are not (an error the certifying agency might have to accept to stay on safe legal grounds with established cut-off scores, but which has possible negative implications for the education of children), and (2) persons classified as "not yet competent" that actually are (which implies denying certification to teachers possibly competent to teach).

Ideally from this model, probabilistic statements would be derived from criterion-related validity coefficients themselves. For example, given a level of assessed teacher performance, what associated yield in pupil learning might be expected. What is the minimally acceptable yield in pupil learning gain desired as a standard, and what teacher performance score (competency) has been associated with it; with what degree of certainty? Given the mixed results of process/product research during the past 15 years, it seems doubtful that measurement technology can make this "productivity function" model useful for certification decisions in the near future.

A second approach to determining cut-off scores for certification can be labelled "criterion-referenced." As used here, this approach would compare assessment data on a particular teacher to an acceptable standard of "minimal" performance established by professional educators as a group. The criteria by which competence is judged is "expert" opinion. Panels of experts

(such as those in studies with the TPAI) would establish scores which in their view represented minimal criteria for entry level performance in the teaching profession. This model is less "statistically" sophisticated than the probabilistic model described above, but it has certain advantages in terms of being easy to implement, apparently on solid legal ground, and better understood by the general public. Major problems arise in contrasting "desired" with "actual" performance levels. If performance levels are set too low, all beginning teachers might automatically pass, thus negating the need for the assessment scheme. Similarly, if performance standards are set too high, large percentages of applicants might fail, placing a heavy burden for staff development and inservice training on the State and local school systems, and raising questions about the quality of higher education institutions in meeting current performance "expectations" through their education programs. A combination of "empirical" and judgemental models for such decisions seems needed.

Legal Issues

The legal issues surrounding the implementation of performance-based certification (PBC) models have their historical basis in Guidelines for fair employment tests and procedures established by the Equal Employment Opportunity Commission (EEOC) as part of Title VII of the Civil Rights Act of 1964. Since 1964, legal, social policy, and measurement issues have pervaded the competency-based education (CBE) and PBC concepts no matter to whom they are applied. No attempt will be made here to review the many court cases derived from the EEOC guidelines. As regards the TPAI, this has been done elsewhere (Johnson, Adams, Miller, Ellett and Capie, 1978). However, issues will be briefly summarized as they have emerged since 1967 (Griggs vs. Duke Power).

Two basic perspectives pervade legal issues surrounding CBE and PBC and past decisions rendered by the courts.

Perspective I: Who says we should have CBE and PBC? Most attempts by states to implement CBE models, whether for teachers, administrators, or children, have their legal basis in state constitutional provisions installed by legislatures as a means of implementing greater educational accountability at all levels.

Perspective II: Who says we should not have CBE and PBC? Most challenges to the CBE and PBC model have come from court decisions rendered against such systems because they are discriminatory, culturally biased, or stigmatizing on the basis of race, social class, sex, age, etc., with the burden of proof on those implementing such systems.

If one were to review the variety of court cases relevant to the CBE movement, especially those involving application of tests or assessment systems for employee selection, certification and licensing, several aspects of the CBE movement have raised important issues. Each of these issues can be the basis for legal challenge. Some of the key issues are summarized below.

Selection of the Competencies

1. Was selection arbitrary?
2. Have competencies been "taught" during professional preparation of the person being assessed?
3. Are the competencies generic and applicable to all persons concerned, e.g., elementary and secondary teachers, teachers in various subject matter areas.

Competency Measuring Devices

1. Were they properly constructed by qualified professionals?
2. Are items on which the person is assessed professionally endorsed?
3. Are measures valid in the sense that they assess job-related content and skills?
4. Are the criteria for validation themselves valid and reliable?
5. Who selected the validation criteria and why?

6. Was an attempt made to have the competency assessment device come up to APA/AERA standards?

Time Sequence of Assessment

1. When will the competency assessment instrument be given?

2. How many times will it be given?

Opportunities for Remediation of Person Assessed

1. Have opportunities for remediation activities for those undergoing and failing an assessment been provided?

2. Do the remediation activities have a rational relationship to both test performance and job-related skills?

3. Are multiple opportunities to "take the test" provided?

Cut-Off Scores

1. How were cut-off scores or competency standards set?

2. Were standards established on an empirical or on an arbitrary bases?

3. Were standards established free of socio-political, fiscal, and other bases?

4. Are standards non-discriminatory?

Due Process Considerations

1. Has a right to a hearing been provided before denial of certification or licensing?

2. Was the right to such hearings guaranteed?

3. Is a failure to be certified free of social stigmatization for the person tested?

4. Are there any adverse impacts on individuals as a result of not passing?

Each of the legal issues raised above in the form of questions has implications for those developing competency tests, exams, and other assessment systems. Most are inextricably interrelated to measurement processes used in instrument development and field testing. Those interested in developing systems to be used for certification like the TPAI would be well advised to plan research and test development activities after first having familiarized themselves with important court decisions. The R & D model is obviously one of making empirical evidence support legal opinion. Appendix D provides some

general recommendations for developing and using performance evaluation measures in the educational milieu adapted from Holley and Feild (1977). Those attempting the development of assessment instruments like the TPAI would be well-advised to design developmental processes and research activities in full view of legal requirements and recommendations.

Summary

This paper presents an overview of a performance-based assessment system developed in the state of Georgia to measure competencies of beginning teachers for initial professional certification. This system (Teacher Performance Assessment Instruments) will be used for the first time during the 1980-81 school year as part of the state's Performance-Based Certification (PBC) plan. To the author's knowledge, the state of Georgia is the first state to commit itself to the complete development and implementation of a PBC model for beginning teachers. Issues related to measurement problems, the establishment of performance standards (cut-off scores), and the legality of performance assessment systems for certification were discussed in view of developmental work surrounding the TPAI. Some of the developmental procedures, problems and associated issues are presented here at a time when many states (particularly those in the Southeast) are contemplating similar performance-based certification models.

References

Berk, R. A. Determination of optimal cutting scores in criterion-referenced measurement. Journal of Experimental Education, 1976, 45, 4-9.

Burke, J. B., Hansen, J. H., Houston, R. W., & Johnson, C. Criteria for describing and assessing competency based programs. Albany, N.Y.: The National Consortium on Competency Based Education Centers, 1976.

Capie, W., Ellett, C. D., & Johnson, C. E. Selected investigations of the reliability of the TPAI. Technical report, Teacher Assessment Project, College of Education, University of Georgia, 1979.

Capie, W., Ellett, C. D., & Johnson, E. E. Relating pupil achievement gains to ratings of secondary student teachers' performance. Paper presented at the annual meeting of the Eastern Educational Research Association, Norfolk, 1980.

Ellett, C. D., Johnson, E. E., Capie, W., & Okey, J. R. A study of the criterion-related validity of the Teacher Performance Assessment Instruments. Research report, Teacher Assessment Project, College of Education, University of Georgia, 1978.

Ellett, C. D., Capie, W., & Johnson, C. E. Teacher performance and pupil achievement on teacher-made tests. Paper presented at the annual meeting of the Eastern Educational Research Association, Norfolk, February, 1980.

Ellett, C. D., Capie, W., & Johnson, C. E. Teacher performance and Pupil perceptions of the social environment of learning. Paper presented at the annual meeting of the Eastern Educational Research Association, Norfolk, February, 1980.

Hambleton, R. K., Powell, S., & Eignor, D. R. A practitioner's guide to criterion-referenced test development, validation, and test score usage. Units from Laboratory of Psychometric and Evaluative Research Report No. 70. Amherst, Ma.: University of Massachusetts, 1979.

Holley, W. H. & Feild, H. S. The law and performance evaluation in education: A review of court cases and implications for use. Journal of Law and Education, 1977, 6, 429.

Houston, W. R., & Howsam, R. B. (Eds.), Competency-based teacher education. Chicago: Science Research Associates, Inc., 1972.

Johnson, C. E., Adams, P. T., Miller, K. A., Ellett, C. D., & Capie, W. A preliminary investigation of the legal validity of the Teacher Performance Assessment Instruments with reviews of court cases. Technical report, Teacher Assessment Project, College of Education, University of Georgia, 1978.

Johnson, C. E., Capie, W. & Ellett, C. D. Experts' opinions of the 1977-78 Teacher Performance Assessment Instruments. Technical report, Teacher Assessment Project, College of Education, University of Georgia, 1978.

Johnson, C., Okey, J., Capie, W., Ellett, C. D., & Adams, P. Identifying and verifying generic teacher competencies. Research Monograph, Teacher Assessment Project, College of Education, University of Georgia, 1978.

Okey, J. R., Capie, W., Ellett, C. D., & Johnson, C. E. Teacher performance validation studies. Technical report, Teacher Assessment Project, College of Education, University of Georgia, 1978.

Quirk, T. J., Witten, B. J. & Weinberg, S. F. Review of studies of the concurrent and predictive validity of the National Teacher Examinations. Review of Educational Research, 1973, Vol. 43, No. 1, 89-113.

Tobin, K., Capie, W., Ellett, C. D., & Johnson, C. E. A confirmatory analysis of the factor structure of teacher performance. Paper presented at the annual meeting of the Eastern Educational Research Association, Norfolk, 1980.

Walberg, H. J., & Anderson, G. J. A manual for the Learning Environment Inventory and the My Class Inventory. Halifax, Nova Scotia, Canda: Atlantic Institute of Education, 1973.

Appendix A

Competency: Communicates with learners both orally and in writing.

Indicator: Gives clear directions and explanations.

Descriptors:

1. Fails to give any directions or explanations (either written or oral) when there is an obvious need to do so.
2. Directions or explanations are hard to understand. Many learners appear confused. No attempt is made to remedy the confusion.
3. The teacher attempts to remedy confusion in directions or explanations. Only a few learners then appear to misunderstand.
4. The teacher identifies specific learners, if any, that have difficulty with directions and explanations and helps them individually.
5. No evidence of learner confusion about directions or explanations is evident or learner problems with directions and explanations have all been satisfactorily handled.

APPENDIX B

Content Outline for the Teacher Performance Assessment Instruments

TEACHING PLANS AND MATERIALS INSTRUMENT

Teaching Skill	Indicators for Assessing Teaching Skills
I. Plans instruction to achieve selected objectives.	1. Specifies or selects learner objectives for lessons.
	2. Specifies or selects teaching procedures for lessons.
	3. Specifies or selects content, materials, and media for lessons.
	4. Specifies or selects materials and procedures for assessing learner progress on the objectives.
	5. Plans instruction at a variety of levels.
II. Organizes instruction to take into account individual differences among learners.	6. Organizes instruction to take into account differences among learners in their capabilities.
	7. Organizes instruction to take into account differences among learners in their learning styles.
	8. Organizes instruction to take into account differences among learners in their rates of learning.
III. Obtains and uses information about the needs and progress of individual learners.	9. Uses teacher-made or teacher-selected evaluation materials or procedures to obtain information about learner progress.
	10. Communicates with individual learners about their needs and progress.
IV. Refers learners with special problems to specialists.	11. Obtains and uses information about learners from cumulative records.
	12. Identifies learners who require the assistance of specialists.
	13. Obtains and uses information from co-workers and parents to assist with specific learner problems.
V. Obtains and uses information about the effectiveness of instruction to revise it when necessary.	14. Obtains information on the effectiveness of instruction.
	15. Revises instruction as needed using evaluation results and observation data.

CLASSROOM PROCEDURES INSTRUMENT

VI. Uses instructional techniques, methods, and media related to the objectives.	1. Uses teaching methods appropriate for objectives, learners and environment.
	2. Uses instructional equipment and other instructional aids.
	3. Uses instructional materials that provide learners with appropriate practice on objectives.
VII. Communicates with learners.	4. Gives directions and explanations related to lesson content.
	5. Clarifies directions and explanations when learners misunderstand lesson content.
	6. Uses responses and questions from learners in teaching.
	7. Provides feedback to learners throughout the lesson.
	8. Uses acceptable written and oral expression with learners.

Appendix B (Cont.)

Content Outline for the Teacher Performance Assessment Instruments

VIII. Demonstrates a repertoire of teaching methods.	9. Implements learning activities in a logical sequence.
	10. Demonstrates ability to conduct lessons using a variety of teaching methods.
	11. Demonstrates ability to work with individuals, small groups, and large groups.
IX. Reinforces and encourages learner involvement in instruction.	12. Uses procedures which get learners initially involved in lessons.
	13. Provides learners with opportunities for participating.
	14. Maintains learner involvement in lessons.
	15. Reinforces and encourages the efforts of learners to maintain involvement.
X. Demonstrates an understanding of the school subject being taught.	16. Helps learners recognize the purpose and importance of topics or activities.
	17. Demonstrates knowledge in the subject area.
XI. Organizes time, space, materials, and equipment for instruction.	18. Attends to routine tasks.
	19. Uses instructional time effectively.
	20. Provides a learning environment that is attractive and orderly.

INTERPERSONAL SKILLS INSTRUMENT

XII. Demonstrates enthusiasm for teaching and learning and the subject being taught.	1. Communicates personal enthusiasm.
	2. Stimulates learner interest.
	3. Conveys the impression of knowing what to do and how to do it.
XIII. Helps learners develop positive self-concepts.	4. Demonstrates warmth and friendliness.
	5. Demonstrates sensitivity to the needs and feelings of learners.
	6. Demonstrates patience, empathy and understanding.
XIV. Manages classroom interactions.	7. Provides feedback to learners about their behavior.
	8. Promotes confortable interpersonal relationships.
	9. Maintains appropriate classroom behavior.
	10. Manages disruptive behavior among learners.

PROFESSIONAL STANDARDS INSTRUMENT

XV. Meets professional responsibilities.	1. Works cooperatively with colleagues, administrators, and community members.
	2. Follows the policies and procedures of the school district.
	3. Demonstrates ethical behavior.
	4. Performs extra-instructional duties.
XVI. Engages in professional self-development.	5. Participates in professional growth activities.
	6. Shares and seeks professional materials and ideas.

Appendix C

Selected Pages from a Computer-Produced TPAI Performance Profile

SUMMARY SHEET OF TEACHING PERFORMANCE

TEACHING PLANS AND MATERIALS INSTRUMENT		RATED COMPETENCY MEAN	RECOMMENDED COMPETENCY MEAN	PLOT OF COMPETENCY MEANS
TPM I:	PLANS INSTRUCTION TO ACHIEVE SELECTED OBJECTIVES	4.33	3.60	 x
TPM II:	ORGANIZES INSTRUCTION TO TAKE INTO ACCOUNT INDIVIDUAL DIFFERENCES AMONG LEARNERS	3.22	3.67	x. . . . .
TPM III:	OBTAINS AND USES INFORMATION ABOUT THE NEEDS AND PROGRESS OF INDIVIDUAL LEARNERS	3.83	3.50	. . x
TPM IV:	REFERS LEARNERS WITH SPECIAL PROBLEMS TO SPECIALISTS	4.44	3.00	. . . . . *
TPM V:	OBTAINS AND USES INFORMATION ABOUT THE EFFECTIVENESS OF INSTRUCTION TO REVISE IT WHEN NECESSARY	3.67	3.50	. x
CLASSROOM PROCEDURES INSTRUMENT				
CP VI:	USES INSTRUCTIONAL TECHNIQUES, METHODS, AND MEDIA RELATED TO THE OBJECTIVES	3.67	3.67	x
CP VII:	COMMUNICATES WITH LEARNERS	4.07	4.20	x
CP VIII:	DEMONSTRATES A REPERTOIRE OF TEACHING METHODS	3.78	3.67	x
CP IX:	REINFORCES AND ENCOURAGES LEARNER INVOLVEMENT IN INSTRUCTION	3.92	4.00	x
CP X:	DEMONSTRATES AN UNDERSTANDING OF THE SCHOOL SUBJECT BEING TAUGHT	3.67	4.00	x . .
CP XI:	ORGANIZES TIME, SPACE, MATERIALS, AND EQUIPMENT FOR INSTRUCTION	3.78	4.00	x .
INTERPERSONAL SKILLS INSTRUMENT				
IS XII:	DEMONSTRATES ENTHUSIASM FOR TEACHING AND LEARNING AND THE SUBJECT BEING TAUGHT	3.78	4.00	x .
IS XIII:	HELPS LEARNERS DEVELOP POSITIVE SELF-CONCEPTS	4.56	4.00	 x
IS XIV:	MANAGES CLASSROOM INTERACTIONS	3.67	4.50	x . . . . .

*RATED COMPETENCY MEAN SCORES GREATER THAN 1 UNIT ABOVE OR BELOW THE RECOMMENDED COMPETENCY MEAN SCORE ARE NOT PLOTTED

Appendix C (Cont.)

Selected Pages from a Computer-Produced TPAI Performance Profile

SUMMARY OF INDICATOR RATINGS FOR EACH COMPETENCY

TEACHING PLANS AND MATERIALS		ACTUAL RATINGS	ACTUAL SUMS	RECOMMENDED RATING	RECOMMENDED SUMS
TPM I: PLANS INSTRUCTION TO ACHIEVE SELECTED OBJECTIVES					
INDICATOR 1.	SPECIFIES OR SELECTS LEARNER OBJECTIVES FOR LESSONS.	5 5 3	13	4	12
INDICATOR 2.	SPECIFIES OR SELECTS TEACHING PROCEDURES FOR LESSONS.	5 5 3	13	4	12
INDICATOR 3.	SPECIFIES OR SELECTS CONTENT, MATERIALS, AND MEDIA FOR LESSONS.	4 5 5	14	3	9
INDICATOR 4.	SPECIFIES OR SELECTS MATERIALS AND PROCEDURES FOR ASSESSING LEARNER PROGRESS ON THE OBJECTIVES.	4 4 3	11	3	9
INDICATOR 5.	PLANS INSTRUCTION AT A VARIETY OF LEVELS.	5 4 5	14	4	12
		TOTAL:	65	TOTAL:	54
TPM II: ORGANIZES INSTRUCTION TO TAKE INTO ACCOUNT INDIVIDUAL DIFFERENCES AMONG LEARNERS					
INDICATOR 6.	ORGANIZES INSTRUCTION TO TAKE INTO ACCOUNT DIFFERENCES AMONG LEARNERS IN THEIR CAPABILITIES.	4 2 1	7	4	12
INDICATOR 7.	ORGANIZES INSTRUCTION TO TAKE INTO ACCOUNT DIFFERENCES AMONG LEARNERS IN THEIR LEARNING STYLES.	4 3 5	12	3	9
INDICATOR 8.	ORGANIZES INSTRUCTION TO TAKE INTO ACCOUNT DIFFERENCES AMONG LEARNERS IN THEIR RATES OF LEARNING.	3 3 4	10	4	12
		TOTAL:	29	TOTAL:	33
TPM III: OBTAINS AND USES INFORMATION ABOUT THE NEEDS AND PROGRESS OF INDIVIDUAL LEARNERS					
INDICATOR 9.	USES TEACHER-MADE OR TEACHER-SELECTED EVALUATION MATERIALS OR PROCEDURES TO OBTAIN INFORMATION ABOUT LEARNER PROGRESS.	3 4 3	10	3	9
INDICATOR 10.	COMMUNICATES WITH INDIVIDUAL LEARNER ABOUT THEIR NEEDS AND PROGRESS.	5 5 3	13	4	12
		TOTAL:	23	TOTAL:	21
TPM IV: REFERS LEARNERS WITH SPECIAL PROBLEMS TO SPECIALISTS					
INDICATOR 11.	OBTAINS AND USES INFORMATION ABOUT LEARNERS FROM CUMULATIVE RECORDS.	2 5 4	11	3	9
INDICATOR 12.	IDENTIFIES LEARNERS WHO REQUIRE THE ASSISTANCE OF SPECIALISTS.	5 5 4	14	3	9
INDICATOR 13.	OBTAINS AND USES INFORMATION FROM CO-WORKERS AND PARENTS TO ASSIST WITH SPECIFIC LEARNER PROBLEMS.	5 5 5	15	3	9
		TOTAL:	40	TOTAL:	27

Appendix D

Recommendations for Developing and Using Performance Evaluation Systems in the Educational Milieu which could help bring them into Compliance with the Law[1]

1. The performance evaluation method should be shown to be job-related through content and/or empirical validation procedures.
2. The content of a performance evaluation method should be developed from a thorough job analysis.
3. Individuals evaluating job performance should frequently observe employees at work.
4. Where possible, evaluations should be based on observable job behaviors.
5. Evaluations should be administered and scored under standardized conditions.
6. Evaluations should not be discriminatory in intent, application, or results.
7. The language, including directions for use, on evaluation forms should be precise and uniform.
8. Evaluators should be trained in how to use the evaluation instruments.
9. Several evaluators should be used, and their evaluations should be conducted independently.
10. Evaluations should be supported by objective evidence of performance results.
11. In selecting factors to be used in performance evaluation or designing evaluation instruments, inputs should be solicited from various groups, for example, faculty members and administrators.
12. Performance evaluation must be conducted before any personnel decisions on which it is based should be made.
13. Criteria and procedures to be used in the evaluations should be communicated to the persons being evaluated.
14. Teachers and faculty members should be informed of the results of their performance evaluation.
15. School systems should develop policies pertaining to the use of performance evaluation, procedures for administration, rules governing decisions, etc., and these should be communicated to the teachers and faculty members.

---

[1] Adapted from William H. Holley and Hubert S. Feild, "The Law and Performance Evaluation in Education: A Review of Court Cases and Implications for Use," Journal of Law and Education 6 (October 1977): 429.